Week 1 Lecture 1

What is data?

Data cannot make decisions – this is where statistics come in

Statistics is generalizing from specific to general

Metadata

Data vs information

Five Ws

Table – what do rows and columns represent?

“spreadsheets”

Relational databases – sql, quick example, show related tables. How could this go wrong? (customer table with location in column)

Sql is usually step 1, we’ll start this course at step 2…

Types of data (categorical and quantitative) – show a table

Quantitative data need units

Are all numbers quantitative?

Identifier variables (student IDs in Banner)

Boolean (not in book?)

Others:

Ordinal (lichert scale) vs nominal (teen, adult, senior)

Time series/ Cross sectional (describe entire dataset, not individual variables)

Data sources: where how and when

Random internet surveys are worthless

Event data example? Who knows what events take attendance or how it’s taken

Experiments – you control how data is collected and intentionally manipulate variables (a/b testing)

Publicly available data – census, WHO, nyc open data